# Package 'crblocks'

April 2, 2019

**Version** 1.0-0

**Date** 2019-03-27

**Title** Categorical Randomized Block Data Analysis

**Author** David Allingham, D.J. Best

**Maintainer** David Allingham <David.Allingham@newcastle.edu.au>

**Description** Implements a statistical test for comparing bar plots or histograms of categorical data derived from a randomized block repeated measures layout.

**License** GPL-3

**URL** https://carma.newcastle.edu.au/davida/

**Imports** stats, utils

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2019-04-02 07:50:09 UTC

## R topics documented:



---

`crblocks` *Categorical Randomized Block Data Analysis*

---

### Description

Implements a statistical test for comparing barplots or histograms of categorical data derived from a randomized block repeated measures layout.

## Usage

```
catrandstat(rawdata)
catrandpvalue(datafilename,Nrepeats)
catrandpvaluepermute(datafilename,Nrepeats)
## S3 method for class 'crblocks_output'
print(x,...)
```

## Arguments

| | |
|---|---|
| rawdata | the data to analyse. |
| datafilename | a character string giving the name of the data file to analyse. |
| Nrepeats | the number of Monte Carlo simulated data sets to use in computing the p-value (10000+ recommended). |
| x | output from catrandstat, catrandpvalue or catrandpvaluepermute |
| ... | not used |

## Details

This package implements the statistical test for comparing barplots or histograms of categorical data derived from a randomized block repeated measures design described in the paper "A Statistical Test for Categorical Randomized Block Sensory Evaluation Data" by DJ Best, JCW Rayner and David Allingham (submitted, 2012). The main functions are `catrandpvalue` and `catrandpvaluepermute`. They read a dataset from a plain-text file can return a p-value, as well as other values of interest, using Monte Carlo simulations and permutations, respectively. The function which computes the statistic can be called directly if desired.

**Data format:**

Using one line of data per judge, each line of the input file contains the category into which each product was placed by that judge, with one column for each product. Each judge must categorise every product.

Comments (starting with #) are allowed (both on their own lines and at the end of lines of data). The file should not contain a header of column names: use a comment to include such descriptions.

There are no error checks on the format. Users should examine the values of Njudges and Nproducts in the output to ensure that they are as expected.

**Note about singular covariance matrices:**

If the covariance matrix of the data is too close to singular, catrandpvalue() can take a very long time to generate the requested number ($Nrepeats) of Monte Carlo data sets. If the number of tries, $Ngenerated, exceeds 1000 $Nrepeats, the simulation is abandoned. In this case, catrandpvaluepermute() should be used, and the appropriate command, with the previously supplied inputs, will be shown.

## Value

For the `catrandstat` function:

| | |
|---|---|
| $Njudges | the number of judges in the data file (number of data lines). |
| $Nproducts | the number of products tested (number of data columns). |

| | |
|---|---|
| $rawdata | a matrix containing the data that was read from the input file (categories for each product by each judge). |
| $categories | a vector containing a list of the categories present in the data. |
| $Ncategories | the number of different categories present in the data (length of $categories). |
| $catCounts | a matrix containing the number of times each product was placed in each category. |
| $judgeCatCounts | a matrix containing the number of times each judge used each category. |
| $Sstatistic | the S statistic computed for the data. |
| $Mstatistic | the M statistic computed for the data. |
| $L2statistic | the L^2 statistic computed for the data. |
| $Schi2pvalue | the chi^2 p-value of the S statistic for the data. |
| $Mchi2pvalue | the chi^2 p-value of the M statistic for the data. |
| $L2chi2pvalue | the chi^2 p-value of the L^2 statistic for the data. |

For the catrandpvalue function:

| | |
|---|---|
| $rawdata | a matrix containing the data that was read from the input file (categories for each product by each judge). |
| $Nproducts | the number of products tested (number of data columns). |
| $Ncategories | the number of different categories present in the data (length of $categories). |
| $Njudges | the number of judges in the data file (number of data lines). |
| $Ngenerated | the number of Monte Carlo data sets generated in total to produce Nrepeats data sets with no ties (where a judge places all products into the same category). |
| $Sdata | the S statistic computed for the data. |
| $Mdata | the M statistic computed for the data. |
| $L2data | the L^2 statistic computed for the data. |
| $Smontecarlo | a vector containing the S statistic values computed for each Monte Carlo data set. |
| $Mmontecarlo | a vector containing the M statistic values computed for each Monte Carlo data set. |
| $L2montecarlo | a vector containing the L^2 statistic values computed for each Monte Carlo data set. |
| $Spvalue | the Monte Carlo p-value for the null hypothesis that there exist no pairwise differences between products based on the S statistic. |
| $Mpvalue | the Monte Carlo p-value for the null hypothesis that there exist no pairwise differences between products based on the M statistic. |
| $L2pvalue | the Monte Carlo p-value for the null hypothesis that there exist no pairwise differences between products based on the L^2 statistic. |
| $Schi2pvalue | the chi^2 p-value of the S statistic for the data. |
| $Mchi2pvalue | the chi^2 p-value of the M statistic for the data. |

| | |
|---|---|
| $L2chi2pvalue | the chi^2 p-value of the L^2 statistic for the data. |

For the catrandpvaluepermute function:

| | |
|---|---|
| $rawdata | a matrix containing the data that was read from the input file (categories for each product by each judge). |
| $Nproducts | the number of products tested (number of data columns). |
| $Ncategories | the number of different categories present in the data (length of $categories). |
| $Njudges | the number of judges in the data file (number of data lines). |
| $Sdata | the S statistic computed for the data. |
| $Mdata | the M statistic computed for the data. |
| $L2data | the L^2 statistic computed for the data. |
| $Spermute | a vector containing the S statistic values computed for each permuted data set. |
| $Mpermute | a vector containing the M statistic values computed for each permuted data set. |
| $L2permute | a vector containing the L^2 statistic values computed for each permuted data set. |
| $Spvalue | the permutation p-value for the null hypothesis that there exist no pairwise differences between products based on the S statistic. |
| $Mpvalue | the permutation p-value for the null hypothesis that there exist no pairwise differences between products based on the M statistic. |
| $L2pvalue | the permutation p-value for the null hypothesis that there exist no pairwise differences between products based on the L^2 statistic. |
| $Schi2pvalue | the chi^2 p-value of the S statistic for the data. |
| $Mchi2pvalue | the chi^2 p-value of the M statistic for the data. |
| $L2chi2pvalue | the chi^2 p-value of the L^2 statistic for the data. |

## Author(s)

Allingham, David <David.Allingham@newcastle.edu.au>

Best, D.J. <John.Best@newcastle.edu.au>

## References

"Comparing Nonparametric Tests of Equality of Means for Randomized Block Designs", Best, D.J., Rayner, J.C.W., Thas, O., de Neve, J., Allingham, D., Communications in Statistics: Simulation and Computation, 45 (5): 1718-1730, 2016.

## Examples

```
 ### Analyse the sample dataset provided with this package:
 # Load the data from the file and compute its test statistic
  inputfile = system.file('extdata', 'omahony.txt', package='crblocks')
  omahonydata=read.table(file(inputfile,'r'))
  closeAllConnections()
  catrandstat(omahonydata)
### OUTPUT:
#
```

```
# Statistic   dof   data value   chi^2 p-value
#  S           6     13.16        0.04058
#  M           2     11.42        0.003311
#  L^2         1     6.671        0.009799
#

 # Load the data from the file and compute the p-value for
 # its test statistic using Monte Carlo simulation:
  catrandpvalue(inputfile,500)
### SAMPLE OUTPUT:
#
# Statistic   dof   data value   chi^2 p-value   Simulated p-value
#  S           6     13.16        0.04058         0.018
#  M           2     11.42        0.003311        0.002
#  L^2         1     6.671        0.009799        0.008
#

 # Load the data from the file, compute the p-value for
 # its test statistic using Monte Carlo simulation, and
 # store the output variables in X:
  Nrepeats = 500
  X = catrandpvalue(inputfile,Nrepeats)
 # This will be a number greater than Nrepeats:
  X$Ngenerated
### SAMPLE OUTPUT:
#
# [1] 6651

 # Load the data from the file and compute the p-value for
 # its test statistic using Monte Carlo simulation:
 catrandpvaluepermute(inputfile,500)
### SAMPLE OUTPUT:
#
# Statistic   dof   data value   chi^2 p-value   Simulated p-value
#  S           6     13.16        0.04058         0.032
#  M           2     11.42        0.003311        0.004
#  L^2         1     6.671        0.009799        0.006
#
```

# Index